Nanopore Targeted Sequencing for the Accurate and Comprehensive Detection of SARS-CoV-2 and Other Respiratory Viruses

Ming Wang, Aisi Fu, Ben Hu, Yongqing Tong, Ran Liu, Zhen Liu, Jiashuang Gu, Bin Xiang, Jianghao Liu, Wen Jiang, Gaigai Shen, Wanxu Zhao, Dong Men, Zixin Deng, Lilei Yu, Wu Wei,* Yan Li,* and Tiangang Liu*

The ongoing global novel coronavirus pneumonia COVID-19 outbreak has engendered numerous cases of infection and death. COVID-19 diagnosis relies upon nucleic acid detection; however, currently recommended methods exhibit high false-negative rates and are unable to identify other respiratory virus infections, thereby resulting in patient misdiagnosis and impeding epidemic containment. Combining the advantages of targeted amplification and long-read, real-time nanopore sequencing, herein, nanopore targeted sequencing (NTS) is developed to detect SARS-CoV-2 and other respiratory viruses simultaneously within 6–10 h, with a limit of detection of ten standard plasmid copies per reaction. Compared with its specificity for five common respiratory viruses, the specificity of NTS for SARS-CoV-2 reaches 100%. Parallel testing with approved real-time reverse transcription-polymerase chain reaction kits for SARS-CoV-2 and NTS using 61 nucleic acid samples from suspected COVID-19 cases show that NTS identifies more infected patients (22/61) as positive, while also effectively monitoring for mutated nucleic acid sequences, categorizing types of SARS-CoV-2, and detecting other respiratory viruses in the test sample. NTS is thus suitable for COVID-19 diagnosis; moreover, this platform can be further extended for diagnosing other viruses and pathogens.

1. Introduction

The novel coronavirus disease (COVID-19) has spread worldwide, resulting in numerous cases of morbidity and death. Generally, COVID-19 has an incubation period of 2–7 days,[1] with no obvious symptoms, during which time the virus can spread from infected to uninfected individuals. Therefore, early accurate diagnosis and isolation of patients is key to controlling the COVID-19 pandemic. Although antibody-based detection methods are rapid, they are readily affected by factors, such as sample hemolysis, the presence of fibrin, bacterial contamination, and patient autoantibodies, resulting in a high false positive rate. Therefore, nucleic acid detection continues to be the gold standard for COVID-19 diagnosis, with several such methods having been employed for detection of the COVID-19 causative virus,

Dr. M. Wang, Prof. Y. Tong, Prof. Y. Li
Department of Clinical Laboratory
Renmin Hospital of Wuhan University
Wuhan 430060, China
E-mail: yanlitf1120@163.com
A. Fu, Dr. B. Hu, Dr. R. Liu, G. Shen, W. Zhao, Prof. Z. Deng, Prof. T. Liu
Key Laboratory of Combinatorial Biosynthesis and Drug Discovery
Ministry of Education and Wuhan University School
of Pharmaceutical Sciences
Wuhan 430071, China
E-mail: liutg@whu.edu.cn

Z. Liu, B. Xiang, Prof. W. Wei
CAS Key Laboratory of Computational Biology
CAS-MPG Partner Institute for Computational Biology
Shanghai Institute of Nutrition and Health
University of Chinese Academy of Sciences
Chinese Academy of Sciences
Shanghai 200031, China
E-mail: wuwei@picb.ac.cn
J. Gu, J. Liu, W. Jiang
Wuhan Dgensee Clinical Laboratory Co., Ltd.
Wuhan 430075, China
Prof. D. Men
Wuhan Institute of Virology
Chinese Academy of Sciences
Wuhan 430071, China
Prof. L. Yu
Department of Internal Medicine
Renmin Hospital of Wuhan University
Wuhan 430060, China
Prof. W. Wei
Center for Biomedical Informatics
Shanghai Engineering Research Center for Big Data in Pediatric
Precision Medicine
Shanghai Children's Hospital
Shanghai Jiao Tong University
Shanghai 200040, China

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

SARS-CoV-2.[2] Specifically, real-time reverse transcription-polymerase chain reaction (RT-qPCR) is currently the most popular testing method for detecting SARS-CoV-2. RT-qPCR is specific, rapid, and economic; however, it is unable to precisely analyze amplified gene fragment nucleic acid sequences. Thus, positive SARS-CoV-2 infection is confirmed by monitoring one or two sites (depending on manufacturer guidelines). Furthermore, RT-qPCR exhibits high false-negative rates in clinical applications,[3] which can facilitate infection transmission through delayed patient isolation and treatment, resulting in continued COVID-19 spread.

Several novel intelligent methods for RNA virus detection have been developed, including combining toehold switch sensors,[4] which can bind to and sense virtually any RNA sequence, using paper-based cell-free protein synthesis. This method has been applied for the detection of Ebola and Zika virus[5,6] and thus should theoretically be capable of rapid and high-throughput SARS-CoV-2 detection as well. Additionally, the SHERLOCK method based on CRISPR/Cas13, can detect Zika, Dengue, and SARS-CoV-2 virus.[7,8] Similarly, the DETECTR method based on CRISPR/Cas12 has been developed for the detection of SARS-CoV-2.[9] Additional new methods based on isothermal PCR amplification are also available, such as Abbott's ID Now instrument, which can interpret results in minutes. However, the requirement for specific RNA regions as targets may negatively affect detection rates as mutation of the target region may limit target availability. Indeed, a 382 nt region of the SARS-CoV-2 genome was found to be deleted in Singapore.[10] Similar deletion events may occur in other regions of the SARS-CoV-2 genome, thereby increasing the risk of acquiring false-negative results if the detection sites are located within the deletion regions.

Sequencing platforms constitute an additional recommended detection method. These platforms are widely applied for pathogen identification and monitoring of virus evolution,[11,12] including that of SARS-CoV-2.[13] Previous massive parallel sequencing platforms sequence DNA by detecting optical or chemical signals. Sequencing by synthesis used in Illumina (the most widely used massive parallel sequencing platform) requires multiple sequencing cycles, each of which takes several minutes to complete, and analyzes a single base of each DNA fragment. Hence the sequencing process generally takes 0.5 to 3 days according to the requirement for read length and data output. Moreover, the sequencing data cannot be applied for further analysis until the entire sequencing process is complete. Nanopore sequencing directly detects changes in currents generated when DNA/RNA molecules pass through a nanopore protein. The speed of DNA/RNA passing through a nanopore protein is incredibly high (≈450 base s^{-1} for DNA and 80 base s^{-1} for RNA). The electrical signal corresponding to each nucleic acid that passes through the nanopore protein can be recorded in real-time and used for subsequent sequence analysis immediately.[14]

The nanopore metagenome method has been shown to effectively detect the respiratory bacterial infection[15] and virus[16,17] directly from clinical samples. Pathogens and antibiotic resistance genes can be identified in several hours, which is much faster compared to traditional culture method as the real-time data generation of nanopore sequencer. Moreover,

nanopore sequencing was used to direct sequences in the transcriptome of SARS-CoV-2, [18,19] as well as the full-length coronavirus genomic RNA.[20] These studies revealed a complex array of viral transcripts with RNA modifications and provided robust estimates of coronaviral evolutionary rates. Alternatively, considering the relatively low abundance of viral nucleic acids compared to that of host nucleic acids in clinical specimens, direct RNA sequencing and metagenome sequencing methods perform unbiased sequence analysis of both viral and human nucleic acids using a substantial amount of sequencing data and thus resulting in exorbitant associated costs and time to complete the analysis. Hence, in most recent studies, ARTIC method, based on tiling multiplex PCR and been used to analyze the Zika[21] and Ebola,[22] was adopted to get the whole genome of SARS-CoV-2 from virus isolates[23] or clinical samples.[24–29] Using this advanced method, by assessing the overlaps among multiplex amplicons, the accurately assembled and complete viral genome can be obtained, which can facilitate the rapid genomic surveillance of SARS-CoV-2 for better understanding its pathogenicity, evolution, and transmission. However, since the large number of clinical samples and the requirement for a short turnaround time, these developed methods are based on nanopore sequencing is not suitable for clinical diagnose and detection of SARS-CoV-2.
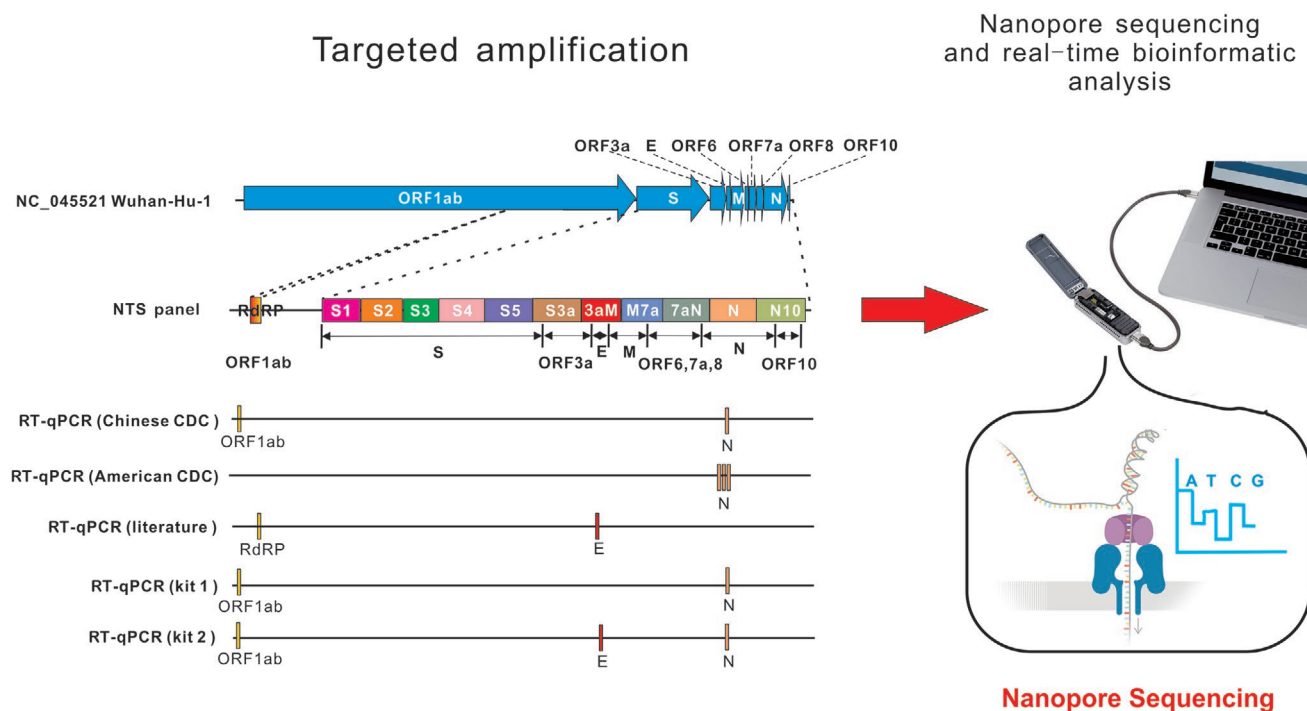
Importantly, pneumonia and fever can also be caused by other respiratory viruses.[30] Cross-infection during the diagnosis process both propagates the spread of SARS-CoV-2 and subjects COVID-19 patients to other respiratory viruses. In severe cases, comprehensive analysis of infecting viruses is necessary. In addition, although thousands of SARS-CoV-2 tests are performed every day around the world, the data obtained by these methods can hardly be used for subsequent analysis of virulence and mutation and for epidemiological investigation. Indeed, nearly all current methods for sequence analysis of the virus are based on whole-genome sequencing methods, which are costly, with low-throughput, thereby limiting the data obtained. However, detection of virulence mutations, virus typing, and epidemiological analysis are critical for the prevention and control of COVID-19. Therefore, a rapid, accurate, and comprehensive detection method is needed to inform clinical treatment and control cross-infection to reduce mortality.

Here, we focus on the diagnosis and detection of SARS-CoV-2 based on nanopore sequencing since mid-January 2020, developed a nanopore targeted sequencing (NTS) platform that can combine the advantages of targeted amplification and long-read, real-time nanopore sequencing with high sensitivity within 6–10 h and simultaneously detect other respiratory viruses, monitor mutated nucleic acid sequences, and categorize types of SARS-CoV-2, and performed NTS for clinical diagnostic tests since early February, 2020.

2. Results and Discussion

2.1. NTS Design for SARS-CoV-2 Detection

Multiplex amplicon sequencing has been proven more sensitive for low-copy SRAS-CoV-2 samples compared to metagenome and capture sequencing, requiring the least amount of

**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

Figure 1. Amplification targets of the NTS and RT-qPCR method. NTS detected 12 fragments including ORF1ab and virulence factor-encoding regions. For RT-qPCR, the Chinese CDC recommends *orf1ab* and *N* sites as targets,[51] the United States CDC recommends three target sites in the *N* gene,[52] and literature recommend RNA-dependent RNA polymerase (RdRP) in *orf1ab* and *E* sites as the targets.[53] Kit 1 is a CFDA-approved kit with two target sites used in this study; kit 2 is a CFDA-approved kit with three target sites used in this study.

sequencing data to identify the virus among the three methods; this indicates that its sequencing time is also the shortest.[31] NTS is based on the amplification of 11 virulence-related gene fragments and one specific gene fragment (*orf1ab*) of SARS-CoV-2 using a primer panel developed in-house, followed by sequencing of the amplified fragments on a nanopore platform. To enhance sensitivity, we focused on virulence-related genes as targets without limitation of the sites currently recommended by Chinese or American Centers for Disease Control (CDC) in RT-qPCR methods (**Figure 1**). Since this method can precisely determine nucleic acid sequences, positive infection can be confirmed by analyzing output sequence identity, coverage, and read number.

To realize the detection of pivotal SARS-CoV-2 virulence genes, we focused on the virulence region (genome bp 21563–29674; NC_045512.2) encoding S (1273 amino acids; AA), ORF3a (275 AA), E (75 AA), M (222 AA), ORF6 (61 AA), ORF7a (121 AA), ORF8 (121 AA), N (419 AA), and ORF10 (38 AA) proteins. We also considered the RNA-dependent RNA polymerase (RdRP) region in ORF1ab (Figure 1). For the virulence regions, 11 fragments of 600–950 bp were designed as targets, providing full coverage of the 9115 bp region (Figure 1). These fragments were amplified by 22 specific primers designed considering primer–primer interactions and annealing temperature and potential nonspecific binding to genomes of human and common bacteria and fungi. To improve the sensitivity of the *orf1ab* region amplification, we designed two primer pairs to amplify 300–500 bp regions to avoid amplification failures owing to site mutation. Finally, the 26 primers were combined to develop the SARS-CoV-2 primer panel (Table S1, Supporting Information).

For sequencing, we chose a nanopore platform capable of sequencing long nucleic acid fragments and simultaneously analyzing the data output in real-time (Figure 1). This allowed for rapid confirmation of SARS-CoV-2 infection by periodical mapping of the SARS-CoV-2 genome sequence reads, as well as analysis of output sequence identity, coverage, and read number. Moreover, the accurate nucleic acid sequence generated using our pipeline effectively indicated whether the virulence-related genes were mutated during virus transmission, thereby rapidly providing information for subsequent epidemiological analysis.

2.2. Turnaround Time, Interpretation, and Limit of Detection (LoD) of NTS

To test the SARS-CoV-2 detection efficiency by NTS, we used standard plasmids harboring COVID-19 virus *S* and *N* genes to simulate SARS-CoV-2. To this end, 0, 10, 100, 500, 1000, and 3000 copies of the standard plasmids were individually spiked into each background cDNA sample (cDNA reverse-transcribed from an uninfected respiratory flora throat swab). These samples then underwent targeted amplification and sequencing performed on a MinION sequencer chip. Sequence data were evaluated at regular intervals using our in-house bioinformatics pipeline. By mapping output reads on the SARS-CoV-2 genome, all reads with >90% identity were

calculated for each plasmid concentration. For 10 min and 1 h sequencing data, reads mapped to SARS-CoV-2 significantly differed from those of negative controls in all replicates at concentrations ranging from 500 to 3000 (**Figure 2**a) and 10 to 3000 (Figure 2b) copies/reaction, respectively. These results confirmed that high-copy samples could rapidly yield sufficient valid sequencing data for diagnosis, and by extending the sequencing time, valid sequencing data could also be obtained from low-copy samples.

As the sequencing time increases, the mapped number of reads in positive samples will also increase; however, the mapped number of reads (1–2 reads) in the negative control (0 copies) will not change significantly (Figure S1, Supporting Information). Therefore, more positive mapped sequencing data should be achieved with additional sequencing time, and clinical samples may exhibit higher complexity; thus, 10 min (for quick detection) and 4 h (for final evaluation) sequencing times were used in the subsequent evaluation of NTS in clinical
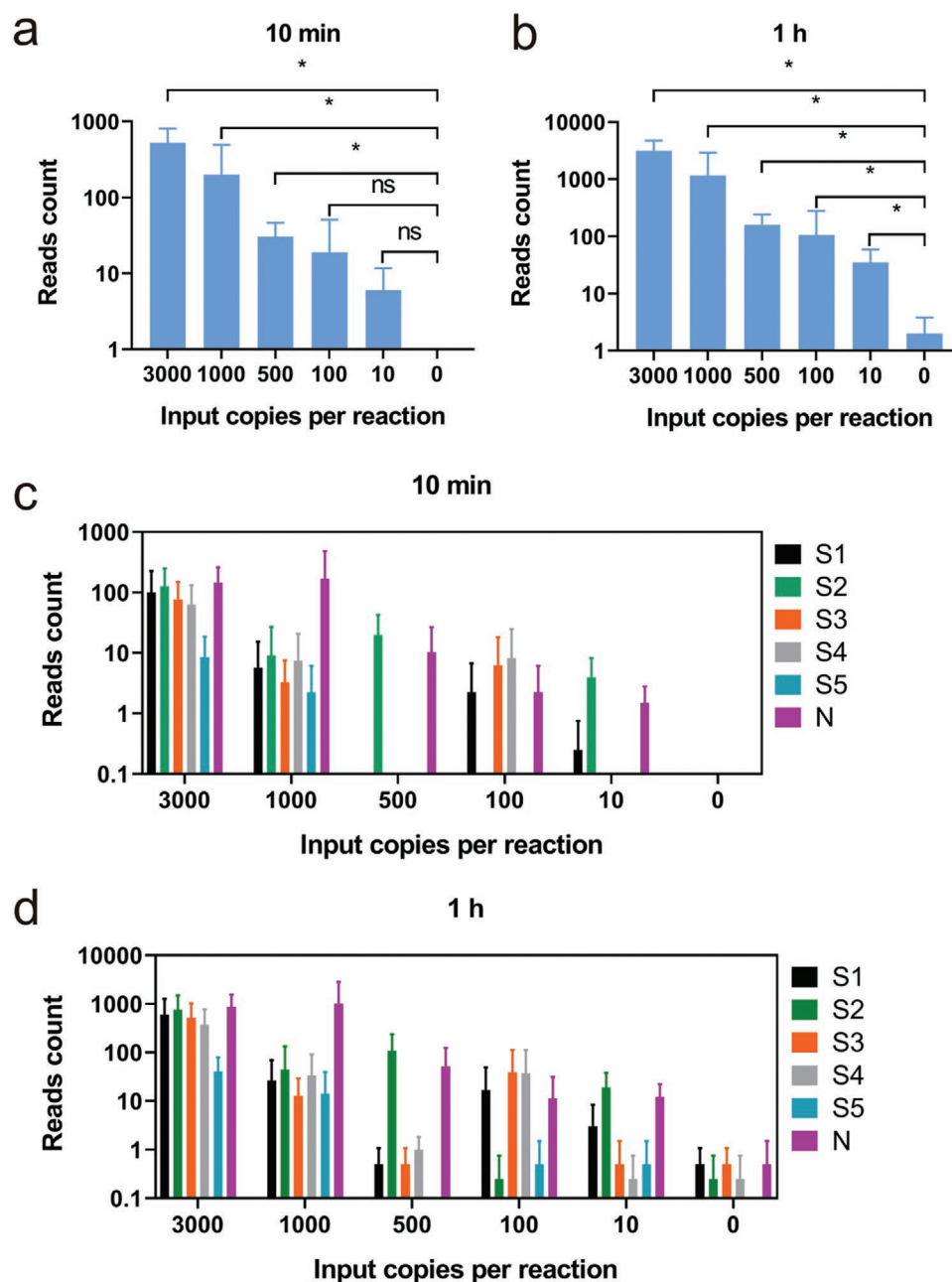
Figure 2. Performance verification test of NTS for detecting SARS-CoV-2 using standard synthetic *S* and *N* genes. Comparison of all SARS-CoV-2 reads detected by NTS in replicates with different concentrations and negative controls using 10 min (a) or 1 h (b) sequencing data. Read counts mapped to each target region of the SARS-CoV-2 genome in replicates with different concentrations with 10 min (c) to 1 h (d) sequencing data. Two-tailed Student t-test (for normal distribution samples) or Mann–Whitney U-test (for non-normal distribution samples): ns, not significant, *$p < 0.05$; bars represent the means ± SD.

samples. The full turnaround time for NTS detection is, therefore, 6–10 h (**Figure 3**), which is longer than that for RT-qPCR; however, 6–10 h is considered acceptable for clinical use. Moreover, NTS is currently the fastest strategy based on sequencing methods for respiratory virus identification to date and can detect sequence variations and virus types directly using clinical samples. It is also important to note that the turnaround time will indeed be prolonged when processing multiple samples manually simultaneously; however, application of the automated operating system that we are currently developing will allow for processing of up to 96 samples at the same time without significantly prolonging the turnaround time.

In the current study, we introduced two rounds of PCR (target PCR and barcoding PCR), for using 96 barcodes to 96 samples that were sequenced on a single chip (Figure 3); this will reduce the cost of each sample test. Of course, a commercial kit (Nanopore native barcode kit, Oxford nanopore technologies, UK) can be used to simplify the procedure of barcoding PCR (the barcode can be directly ligated to PCR products in the process of library preparation); this may reduce the possibility of contamination. However, the kit allowed for a maximum throughput of 24 samples on one chip as only 24 barcodes are provided as yet, which will increase the cost per sample. Importantly, when processing the samples, we conducted nucleic acid extraction, preparation of PCR reaction, and purification of PCR products in different rooms to reduce the possibility of sample cross-contamination.

Evaluation of the target distribution of the datasets revealed that in higher-copy samples (1000 and 3000 copies per reaction), all targeted regions were detected (Figure 2c,d). However, in lower-copy samples (10–500 copies per reaction), some of the targeted regions were lost (i.e., no reads mapped; Figure 2c,d), indicating that for low-quality or low-abundance samples, comprehensive fragment amplification is difficult. Therefore, for accurate results, NTS cannot label a sample as positive for infection by monitoring only one or two sites, as is customary for RT-qPCR; rather, the results from all target regions should be considered.

In fact, although NTS combines targeted amplification and sequencing, its judgment is more similar to sequencing than RT-qPCR. Hence, we determined a scoring rule by referring to previous judgment rules.[32–34] Firstly, we counted the number of output reads with >90% identity to the SARS-CoV-2 genome and the read matches in a region within 50 bp upstream of the start and 50 bp downstream of the end of the design fragment, which was indicative of high credibility of identification as SARS-CoV-2. By calculating the ratio of the counted valid read numbers of the test sample to those of the negative control (with "0" in the negative control calculated as "1" to avoid having no value after multiplying), considering that there may also be 1–2 reads in negative samples that are misjudged as mapped reads, we defined that a high ratio of ≥10 (to eliminate mismatch interference) indicates a positive result for that fragment, scoring 1; that of ≥3 to 10 is inconclusive, scoring 0.4; and that of <3 is negative, scoring 0. Scores were summed to obtain the NTS score. We defined a sample in which at least 50% fragments (six fragments) were inconclusive or two fragments were positive (comparable to RT-qPCR results) to be a positive infected sample (e.g., NTS score >2.4); one that contained 3–6 inconclusive fragments or 1 positive fragment to be a highly suspect (inconclusive) sample (e.g., NTS score of 1.2–2.4); and one that contained <3 inconclusive or no positive fragments to be a negative sample (NTS score <1.2).

According to our scoring system, the highest NTS score for SARS-CoV-2 detection is 12. Alternatively, the standard plasmids contain only six designed fragments (half of 12 designed fragments for SARS-CoV-2), indicating that the highest score for the simulated tests is 6. Therefore, we reduced the score criterion in simulation experiments by 50%; hence, in simulation tests using the standard plasmids, the NTS score >1.2 indicates positive detection, 0.6–1.2 is inconclusive, and <0.6 reflects negative detection. To determine the NTS LoD, we also used a similar rule as that used for LoD determination in metagenomic sequencing,[15] which uses the defined scoring rules to evaluate each replicate in the simulated test; the lowest concentration of the positive control that can be positively detected (3/4 replicates positive) was set as the LoD. We calculated the score of the lowest concentration (ten copies) at different times according to this scoring method and judged the positive detection rate. The
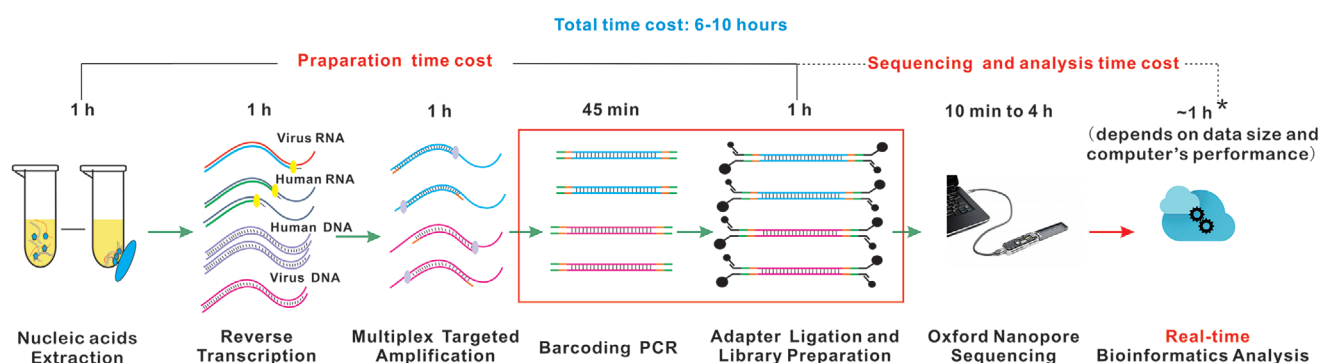
Figure 3. Turnaround time of NTS. The total nucleic acids, including single-stranded DNA/RNA and double-stranded DNA, were extracted, and the total RNA in the total nucleic acids was reverse transcript to cDNA. Specific regions of the DNA virus and cDNA of the RNA virus were amplified by multiplex PCR (one tube for SARS-CoV-2 and another tube for respiratory viruses). Next, the same barcode was added to both ends of the PCR product from the same sample using a barcoding PCR step. The barcoded products of each sample were pooled and used for sequencing library preparation. The barcoding PCR step in the red frame can be removed by directly ligating the barcode to products of multiplex PCR during library preparation using a commercial kit, the turnaround time and risk of cross-contamination could be further reduced. Time for bioinformatics analysis depends on data size and the computer's performance.

**ADVANCED
SCIENCE NEWS**
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

results (Table S2, Supporting Information) showed that 3/4 of the ten copies of the standard plasmids can be judged positive using the 1 h sequencing data. This result is consistent with the significant comparation (Figure 2b) that the data for ten copies of standard plasmids differ significantly from those of the negative control from 1 h. This result shows that our scoring system is reliable for evaluating NTS test results, and the LoD (3/4 replicates positive) was determined as ten copies per reaction with 1 h sequencing data (1372–43967 reads per sample in a run with 24 samples).

2.3. Specificity of NTS for SARS-CoV-2 Detection

To verify the specificity of the SARS-CoV-2 primer panel in NTS, we selected five virus-positive throat samples (influenza A virus, influenza B virus, parainfluenza, respiratory syncytial virus, and rhinovirus), that is all positive samples collected from November 2019 to January 2020 at the Department of Clinical Laboratory, Renmin hospital of Wuhan University, all of which were previously confirmed using a China Food and Drug Administration (CFDA) approved kit (Health Gene Technologies, China) based on multiplex PCR and capillary electrophoresis analysis. These five samples were tested in duplicate by NTS using the SARS-CoV-2 primer panel (**Figure 4**). After a 4 h sequencing, each test sample generated 59 004–156 032 reads, of which, over 99.99% of the reads could not be mapped to any virus genome, and the remainder of the reads mapped to human endogenous retroviruses. The analysis pipeline of NTS can distinguish SARS-CoV-2 with those respiratory viruses.

Since we are unable to collect additional respiratory viruses, the next step is to theoretically analyze the possible match position using the SARS-CoV-2 primer panel to six common human coronaviruses (human coronavirus 229E: NC_002645.1, human coronavirus HKU1: NC_006577.2, human coronavirus OC43: NC_006213.1, human coronavirus NL63: NC_005831.2, SARS: severe acute respiratory syndrome, NC_004718.3, MERS:
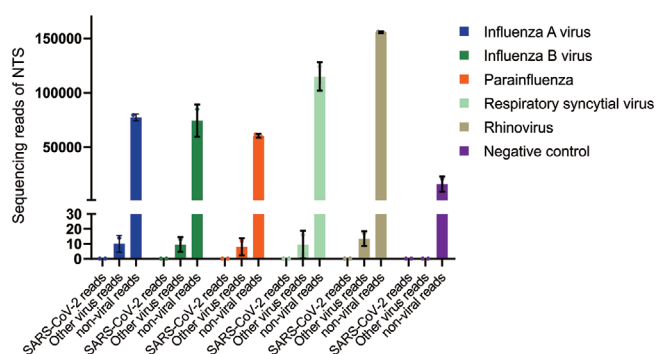
Figure 4. Specificity test. Five throat samples containing influenza A virus, influenza B virus, parainfluenza, respiratory syncytial virus, and rhinovirus were selected to test the cross-reactivity of the SARS-CoV-2 primer panel for common respiratory viruses in duplicate. TE buffer spiked with human DNA was parallelly tested as a negative control. None of the sequencing reads could be correctly mapped to the SARS-CoV-2 genome in all samples and the negative control. Nonviral reads could not correctly be mapped to any reference in the viral genome database, which may derive from the nonspecific amplification of human genome. Several sequencing reads in samples could be mapped to other virus genomes.

Middle East respiratory syndrome, NC_019843.3). The results showed that the SARS-CoV-2 primer panel only matched five positions on the SARS genome and one position on the HKU1 genome (Table S3, Supporting Information), indicating that the primers used (Orf1ab-F1/R1 or Orf1ab-F2/R1) only amplify 400–460 bp fragments of the SARS genome among the six coronaviruses. For the other viruses, fragments were not significantly amplified, indicating that the resulting sequencing data will not include genome fragments of these viruses.

By comparing the similarity between these common coronaviruses and the 12 amplified fragments designed for SARS-CoV-2 amplification, we found (Table S4, Supporting Information) that the identity of RdRP fragments to the SARS genome was 0.916, which is a relatively high identity; however, since 100% homology was not observed, the amplified fragment (RdRP) can be used for distinguishing SARS from SARS-CoV-2 by mapping the reads to all of the virus genomic sequences in the database. Alternatively, if the RdRP fragment of SARS was incorrectly identified as that of SARS-CoV-2, according to the NTS scoring rules, the sample would be assigned a score of 1 and judged as negative. Therefore, among the current coronaviruses, NTS detection of SARS-CoV-2 theoretically has a very high specificity.

2.4. SARS-CoV-2 Detection Using RT-qPCR versus NTS

We performed NTS for 61 throat swab clinical samples collected from 61 patients at the first-line hospital in Wuhan once the NTS method was established (**Figure 5**). The samples were divided into two groups: i) 45 nasopharyngeal swabs from outpatients with suspected COVID-19 early in the epidemic (January 2020), for whom detailed records and suitable clinical data were unavailable, preventing us from confirming SARD-CoV-2 infection. ii) 16 nasopharyngeal swabs from hospitalized patients who had been diagnosed with COVID-19 by clinicians through comprehensive results of nucleic acid tests, chest computed tomography scans, blood tests, and clinical symptoms and, hence, for whom nucleic acid samples tested positive for SARS-CoV-2. The median age of the hospitalized patients was 47 years (range 26–76 years), with seven (44%) males and nine (56%) females.

To test NTS performance, we first evaluated the 45 nasopharyngeal swab samples from outpatients. On February 6 and 7, 2020, we parallel tested these 45 samples in two batches using NTS (two chips) and RT-qPCR (kit 2; Figure 1). The NTS sequencing reads were evaluated to be with high quality and mapping identity (Figure S2, Supporting Information). The 4 h sequencing output data (**Figure 6**a), revealed that all 19 samples, defined as positive by RT-qPCR, were recognized as SARS-CoV-2-infected by NTS, indicating good inter-test concordance. Among 15 RT-qPCR-inconclusive samples, 11 were recognized as SARS-CoV-2-infected, 3 as negative, and 1 inconclusive by NTS. Among 11 RT-qPCR-negative samples, 4 were recognized as SARS-CoV-2-infected, 4 as inconclusive, and 3 as negative by NTS. Overall, NTS identified a total of 34 positive samples among 45 suspected samples, which was 15 more than the number detected by RT-qPCR. Evaluation of output data after 10 min of sequencing (Figure S3, Supporting Information)
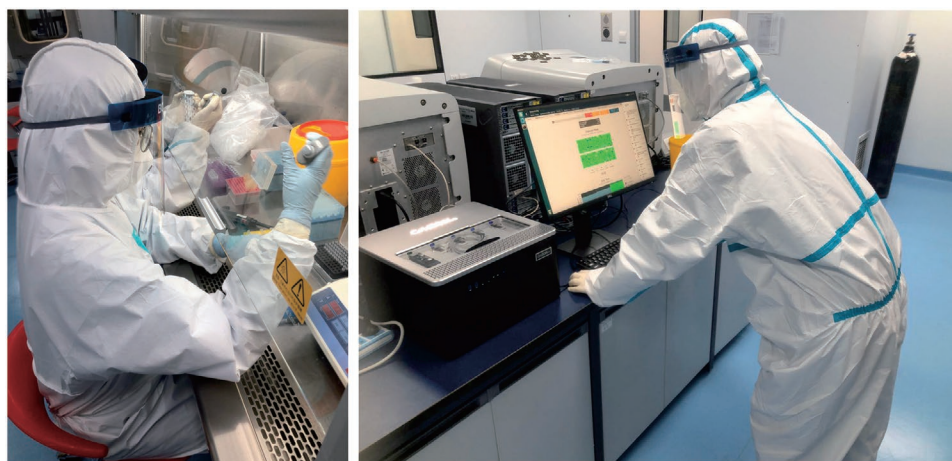
Figure 5. NTS testing in a front-line hospital in Wuhan.

revealed that 21 of 45 suspected samples were recognized as SARS-CoV-2-infected by NTS. For these samples, the 10 min and 4 h sequencing results were comparable, indicating that NTS could rapidly detect many positive samples.

However, as the 45 tested samples were from early outpatients without detailed records, suitable clinical data, such as chest computed tomographic scans, were not available to support the results. Therefore, we next evaluated samples retained from hospitalized patients with confirmed COVID-19 subjected to RT-qPCR testing (kit 1, Figure 1) on February 11 and 12, 2020. We randomly selected 16 patients' samples for NTS testing on February 20, 2020. Following 4 h sequencing (Figure 6b), that sequencing reads were evaluated to be with high quality and alignment identity (Figure S2, Supporting Information), all 16 samples tested positive, whereas only 9 samples were positive by RT-qPCR. At the time of writing this manuscript, among the seven samples that were deemed negative or inconclusive via RT-qPCR, electronic records indicated that subsequent RT-qPCR testing for four of these seven patients revealed two (R04 and R09) as positive and two (R06 and R07) as inconclusive. These results suggest that NTS could identify positive COVID-19 infected cases, seems has higher identification ability than RT-qPCR. Moreover, three positive samples were identified by 10 min sequencing data (Figure S3, Supporting Information), indicating that NTS could rapidly detect positive samples with a high concentration of virus.

Evaluation of the positive target distribution for each sample (Figure 6) indicated that samples positive by both NTS and RT-qPCR had higher nucleic acid quality or abundance, as NTS yielded more positive fragments. For RT-qPCR-inconclusive samples, NTS yielded few, scattered positive target fragments, suggesting that low sample nucleic acid quality or abundance rendered it difficult to draw clear conclusions by RT-qPCR based on evaluation of only two sites. Moreover, the designed amplified fragments are 300–950 bp in length in NTS; these are suitable lengths for detection by a nanopore sequencing platform, as nucleic acid fragments <200 bp cannot be readily detected;[35,36] hence, the sensitivity of NTS for detecting target SARS-CoV-2 fragments using highly degraded nucleic acids may be hampered. The negative control of the first experiment

in Figure 6a appears to have been contaminated with a fragment containing the *N* gene, according to the rule, only the *N* fragment of C1 and H9 were judged as positive and that of another four samples (A4, C2, D12, and G4) were judged as inconclusive. Indeed, the other 11 fragments of these samples were successfully amplified, which means these samples real containing SARS-CoV-2. Moreover, the final sample result (NTS score) was scored according to all 12 fragments, so contamination of an individual SARS-CoV-2 fragment did not affect the final NTS results. However, if random contamination of multiple fragments in negative control occurs, the data in this batch of experiments could not be judged and the experiments need to be repeated.

2.5. SARS-CoV-2 Mutation Analysis and Types Classification

Mutation screening of 50 NTS positive samples following the Medaka variant calling process for haploid genomes, filtered by quality values ≥30 and sequencing depth ≥10, identified a total of 42 single base mutations among 27 samples (Ref. NC_045512.2), 14 of which were synonymous and 28 of which were nonsynonymous mutations (**Table 1**).

Among the 28 nonsynonymous mutations, T28144C (Leu→Ser) occurred eight times, G28077C (Val→Leu) occurred two times, and the remaining nonsynonymous mutations were observed only once. Tang et al. (2020) have found that SARS-CoV-2 genomes evolved into two major types (designated L and S) that are well defined by two different SNPs at position 8782 (T8782C, synonymous) and 28144 (T28144C, Leu→Ser).[37] Based on the classification of 50 NTS positive samples (Experimental Section), 31 samples had a ≥10× depth at position 28144 (Table S5, Supporting Information), of which 22 (71.0%) were classified as L type, 8 (25.8%) were classified as S type, and one (3.2%) was uncertain (**Figure 7**c). These results were consistent with those previously reported,[37] which indicated that the L type was more prevalent in the early stages of the outbreak in Wuhan. Furthermore, genome wide allele frequency analysis (Experimental Section) of 50 NTS positive samples, as well as 1145 recently published (before March 24, 2020) SARS-CoV-2

2002169 (7 of 15)

a

Paitent ID	RdRP	S1	S2	S3	S4	S5	S3a	3aM	M7a	7aN	N	N10	NTS Score	NTS	RT-qPCR
A2	96	60	118	75	40	4	12	14	60	53	132	4	9.8	Pos	Pos
A4	797	793	1079	901	744	64	123	190	439	1179	2014	32	11.4	Pos	Pos
C1	2768	2213	2714	1882	2212	202	620	498	2538	2628	3167	364	12	Pos	Pos
C2	1680	1883	3491	1810	2445	92	737	268	1151	2973	3052	464	11.4	Pos	Pos
D12	3731	2683	4080	2392	3868	464	1709	815	2193	4825	2738	116	11.4	Pos	Pos
E5	124	211	324	121	175	9	70	24	163	359	385	3	9.8	Pos	Pos
G1	2026	791	1636	1375	1158	109	382	252	1335	2532	2921	82	11.4	Pos	Pos
H9	2015	2260	3624	1817	2691	223	647	481	1990	2621	3273	18	12	Pos	Pos
E1	582	237	258	215	180	36	54	19	237	377	360	0	10	Pos	Pos
A6	7	7	9	3	4	3	7	0	1	4	8	1	3.2	Pos	Inc
A9	11	1	40	2	1	7	0	5	20	63	55	0	4.8	Pos	Inc
A12	6	4	5	6	10	1	4	2	4	12	161	0	3.6	Pos	Inc
B4	3	6	11	6	10	1	4	1	2	8	6	1	3.4	Pos	Inc
C11	7	10	5	3	11	1	3	0	1	0	35	1	3.2	Pos	Inc
D2	0	2	8	2	1	0	2	2	4	4	6	1	0.8	Neg	Inc
F5	208	233	3	149	6	0	2	0	2	6	9	1	3.8	Pos	Inc
F12	3	1	5	1	1	1	1	0	0	1	7	0	0.4	Neg	Inc
H12	14	14	16	10	8	2	1	2	10	7	43	2	4.6	Pos	Inc
D10	3	2	5	1	1	0	0	0	1	2	2	0	0.4	Neg	Neg
A1	14	7	14	8	8	4	9	1	6	7	10	0	4.2	Pos	Neg
A3	5	10	13	2	4	1	0	1	3	2	8	0	2.2	Inc	Neg
A5	0	0	2	1	1	0	1	0	0	0	9	2	0	Neg	Neg
A7	1	8	3	1	0	1	1	0	0	22	549	41	2.4	Inc	Neg
A8	2	6	7	2	5	1	0	1	0	0	6	1	1.2	Inc	Neg
PC	0	622	504	386	517	339	26	0	0	0	0	0	6	Pos	Pos
NC	1	0	2	0	0	0	0	0	2	0	315	1	NA	Neg	Neg
A11	150	17	9	67	12	5	6	0	1	810	294	0	6.6	Pos	Pos
B9	1410	179	711	808	198	9	637	75	146	1117	1285	85	11.4	Pos	Pos
C9	1145	190	547	89	421	12	69	48	135	1317	63	0	11	Pos	Pos
D9	702	122	1440	708	769	28	807	45	74	2459	1931	0	11	Pos	Pos
D11	270	324	517	71	418	4	306	32	96	527	604	0	10.4	Pos	Pos
E3	3	2	94	37	132	15	3	0	18	5	133	1	7.2	Pos	Pos
E11	353	277	766	302	231	11	536	40	76	656	1078	157	12	Pos	Pos
F11	2009	511	1307	1166	390	30	1604	129	245	2092	2631	0	11	Pos	Pos
G11	2709	1944	2885	1735	1983	71	2334	418	595	3523	2964	8	11.4	Pos	Pos
G12	4200	1202	4765	1433	1532	6	4131	58	1711	6722	1125	0	10.4	Pos	Pos
C12	125	187	110	2	1	12	99	23	60	88	91	2	9	Pos	Inc
E6	118	123	224	2	16	1	2	16	3	3	2	1	5.8	Pos	Inc
E12	6	6	3	4	3	3	6	0	0	8	6	1	3.2	Pos	Inc
F8	5	157	5	7	295	2	5	0	1	332	239	0	5.2	Pos	Inc
G6	1	4	173	4	7	1	2	0	0	3	2	0	2.2	Inc	Inc
H3	1	0	4	26	1	1	0	0	0	0	0	0	1	Neg	Inc
A10	0	0	0	0	0	0	0	0	0	0	52	0	1	Neg	Neg
B1	1	5	7	3	6	33	3	0	0	4	8	0	3.4	Pos	Neg
B2	27	4	1	2	11	0	0	0	0	15	2	0	3	Pos	Neg
B3	21	76	10	8	118	4	8	0	14	212	288	1	7.6	Pos	Neg
B5	4	6	4	3	9	1	0	0	1	6	4	0	2.4	Inc	Neg
PC	0	916	848	588	848	593	17	0	0	0	1	0	6	Pos	Pos
NC	1	2	2	0	0	0	1	0	0	0	0	0	NA	Neg	Neg

The "NTS target sites" label spans the columns RdRP through N10. "Comparison of NTS and RT-qPCR" spans NTS Score, NTS, RT-qPCR.

b

Paitent ID	RdRP	S1	S2	S3	S4	S5	S3a	3aM	M7a	7aN	N	N10	NTS Score	NTS	RT-qPCR
R01	1	4	38	0	2	0	3	0	0	11	11	0	3.8	Pos	Pos
R02	7	2	12	1	6	0	6	0	2	5	7	1	3	Pos	Neg
R03	50	0	17	1	3	0	10	1	0	13	8	0	4.8	Pos	Pos
R04	1	0	13	1	0	0	3	0	0	5	18	0	2.8	Pos	Neg
R05	135	0	127	23	12	0	176	1	109	580	460	0	8	Pos	Pos
R06	90	14	42	31	1	1	38	0	0	2	95	0	6	Pos	Inc
R07	0	0	4	7	12	2	1	0	0	28	9	0	3.2	Pos	Inc
R08	33	1	4	1	4	0	3	0	1	1	3	0	2.6	Pos	Inc
R09	6	2	8	2	6	0	16	0	0	7	3	1	3	Pos	Neg
R10	3	2	59	0	18	0	4	0	1	4	55	0	4.2	Pos	Neg
R11	6131	496	8613	1310	3526	38	5364	60	317	5481	5436	178	12	Pos	Pos
R12	610	100	882	120	55	9	340	18	2	898	345	0	9.4	Pos	Pos
R13	40	0	0	0	0	2	191	1	0	65	0	0	3	Pos	Pos
R14	7	0	15	1	7	0	6	0	1	6	11	0	3.6	Pos	Pos
R15	35	11	188	0	93	0	65	0	2	75	49	0	7	Pos	Pos
R16	8	0	11	0	6	0	8	0	0	3	6	1	3	Pos	Pos
NC	0	0	0	0	0	0	0	0	0	0	0	0	NA	Neg	Neg
NC	0	0	0	0	0	0	0	0	0	0	0	0	NA	Neg	Neg

Legend: ■ Positive ■ Inconclusive ■ Negative

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

genomes from the GISAID database[38] also indicated a common SNP at position 28144 (Figure 7a), with a similar allele frequency (26.7% of NTS and 28.3% of GISAID). Due to only six samples having ≥10× depth at 25 304 and 29 483, there was only one sample to support the presence of a T25304A or T29483G mutation, and the allele frequency of T25304A and T29483G was 16.7%.

The relationship between mutation sites can be clearly visualized by linkage disequilibrium (LD) plots (Figure 7b and Figure S4, Supporting Information). We, therefore, constructed these plots using filtered Medaka analysis results, while D values were used to represent the linkage of mutation pairs. In the LD plot, most mutation pairs did not exhibit significant linkage. In fact, although the mutations C24034T and G28077C had the highest D value (0.14), only two samples (E5, G11) supported the linkage between these mutations within 27 samples. Therefore, there was not enough evidence to prove that any mutation pairs had a significant linkage.

2.6. NTS Panel for Respiratory Virus Identification

The inability of current, clinically utilized SARS-CoV-2 RT-qPCR kits to identify species of co-infecting viruses, combined with the high false-negative rate of RT-qPCR compromises early patient triage, resulting in wasted urgent medical resources and enhancing potential cross-contamination during the diagnosis process. Hence, distinguishing different types of respiratory viral infections has attracted worldwide attention.

To extend the scope of NTS-based virus detection, we designed a respiratory virus primer panel for amplification of ten respiratory viruses, including bocavirus, rhinovirus, human metapneumovirus, respiratory syncytial virus, coronavirus, adenovirus, parainfluenza virus, influenza A virus, influenza B virus, and influenza C virus. We then collected target gene candidates utilized for virus identification in the literature, and collected all complete and partial target gene sequences for these viruses available in GenBank (through November 1, 2019). Though multiple nucleic acid sequence alignments were available for each gene, the conserved regions were chosen as candidate regions for amplification. Using similar constraints as those applied for SARS-CoV-2 target region selection, we chose 20 target amplification regions (300–800 bp) for the ten respiratory viruses (Table S6, Supporting Information) capable of accurately distinguishing viruses in addition to identifying virus species. We designed 59 primers for the amplification of these regions, comprising the respiratory virus primer panel (Table S7, Supporting Information).

To verify the performance of this panel in NTS, we selected five throat samples positive for five different viruses as mentioned above (influenza A virus, influenza B virus, parainfluenza, respiratory syncytial virus, and rhinovirus). The five samples were mixed to create a mock virus community and used to test the NTS virus detection capacity. NTS 10 min sequencing data (Table S8, Supporting Information) successfully detected four of the five viruses (influenza A virus, influenza B virus, respiratory syncytial virus, and rhinovirus); the remaining virus, with a lower viral load, was detected through 2 h sequencing. As these samples were obtained from an actual clinical setting, they confirmed the suitability of performing NTS with the respiratory virus primer panel, for the clinical identification of at least five kinds of respiratory viruses. Since there were no available positive specimens for the other five viruses during the collection period, we were only able to speculate that they would also be detectable with this panel.

To verify the ability of NTS to detect SARS-CoV-2 and other respiratory viruses within a single assay, 13 of the 45 suspected COVID-19 outpatient samples were subjected to simultaneous detection analysis. Five replications of the plasmid containing the SARS-CoV-2 S and N genes served as the positive control, and Tris-EDTA (TE) buffer was used as the negative control (in duplicate). For each sample, cDNA samples were separately amplified using the respiratory virus and the SARS-CoV-2 primer panels, after which all amplified fragments were pooled. Following the addition of barcodes, amplified fragments from all 20 samples (13 cases, 7 controls) were subjected to nanopore sequencing on one chip. Analysis of the results (**Table 2**) revealed that E11 was co-infected by influenza A virus H3N2 and SARS-CoV-2.
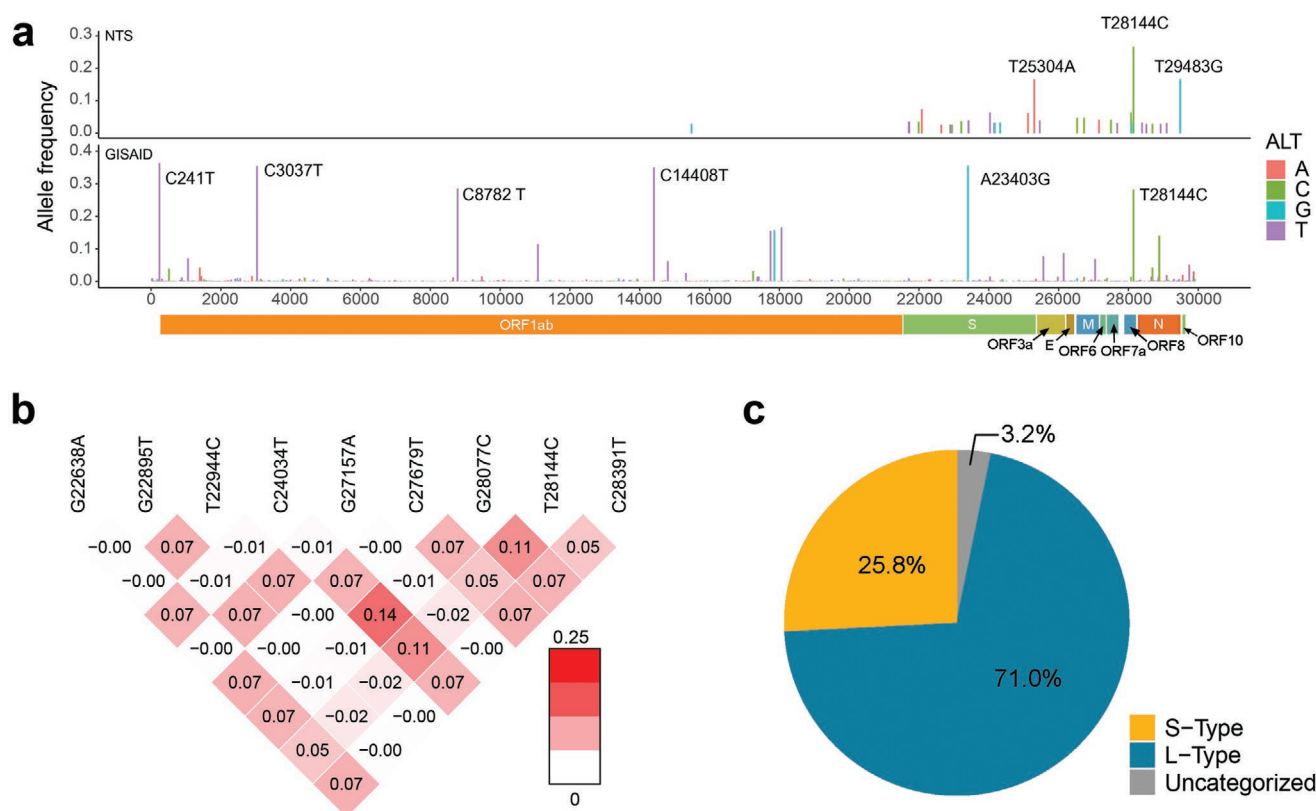
3. Conclusion

Herein, we developed an NTS method capable of simultaneously detecting SARS-CoV-2 and additional respiratory viruses within 6–10 h. Moreover, 22 of the 61 suspected COVID-19 samples that tested negative or inconclusive by RT-qPCR testing, were identified as positive by NTS. This platform also enabled the detection of virus mutations and may be effective for typing of SARS-CoV-2, providing supporting data for future virulence and epidemiological analyses of the virus. However, NTS is not the method that will solve all challenges associated with SARS-CoV-2 detection. In fact, the turnaround time for NTS is longer than that of RT-qPCR, and its operation requires more skill than that of RT-qPCR. Therefore, we believe that NTS and RT-qPCR are complementary platforms as RT-qPCR can rapidly diagnose patients with high nucleic acid content, while NTS can further diagnose patients who cannot be accurately assessed via RT-qPCR. NTS requires further improvements as our current process and the resulting sequencing data analysis and interpretation are not yet mature. Hence additional NTS test results will be collected, and the process will be continuously optimized to obtain more accurate results. In the future, the introduction of integration systems or sealed devices, such as microfluidics, may further avoid

Figure 6. Comparison of 61 nucleic acid samples from clinical samples obtained using NTS (4 h) and RT-qPCR. a) Comparison of 45 nucleic acid samples from samples of patients with suspected COVID-19 obtained using NTS and RT-qPCR (kit 2). b) Comparison of 16 nucleic acid samples from patients with confirmed disease obtained using NTS and RT-qPCR (kit 1). The numbers in the table on the left represent the number of mapped reads according to our rules. PC: positive control. The plasmid harboring an S gene was used as a positive control in NTS testing; a positive sample in the kit served as a positive control in RT-qPCR testing. NC: negative control. TE buffer was used as a negative control in NTS testing; H_2O in the kit served as a negative control in RT-qPCR testing. All positive sample and negative sample in NTS were introduced from nucleic acid extraction. Pos: positive. Inc: inconclusive. Neg: Negative.

Table 1. Variations of SARS-CoV-2 detected by NTS. Reference genome: NC_045512.2. Quality: the quality score of variant output by medaka variant program. Depth: the depth of sequencing reads mapped to genomic position. (Reads contain deletion at the position are not counted). VAF: Variant allele fraction. Reads contain alternative nucleotide/depth.

Patient ID	Genomic position	Reference	Variant	Quality	Depth	VAF	Amino acid change
R13	15484	T	G	56.227	38	0.947368	Ser→Ala
R11	21707	C	T	63.216	475	0.987368	His→Tyr
B3	21721	C	T	58.097	72	0.902778	NC
R06	21992	T	C	39.613	14	0.857143	Tyr→His
D11	22081	G	A	35.992	306	0.735294	NC
C1	22081	G	A	52.609	2091	0.918221	NC
R15	22638	G	A	63.216	186	0.924731	Ser→Asn
R05	22895	G	T	36.94	126	0.730159	Val→Phe
R05	22944	T	C	42.161	124	0.798387	Leu→Pro
F5	23212	T	C	66.227	145	0.944828	NC
B9	23416	A	T	60.786	769	0.886866	NC
G12	23422	C	T	59.237	1314	0.952816	NC
E5	24034	C	T	56.684	153	0.862745	NC
R15	24034	C	T	58.445	89	0.932584	NC
E3	24154	G	A	57.196	122	0.901639	NC
R12	24172	T	G	52.902	55	0.872727	Ile→Met
B9	24325	A	G	36.687	185	0.843243	NC
B1	25124	G	A	61.455	31	0.935484	Glu→Lys
F11	25304	T	A	58.097	28	0.821429	Cys→Ser
E11	25460	C	T	66.227	491	0.940937	Ala→Val
E6	26531	T	C	49.994	14	0.928571	NC
E5	26729	T	C	52.802	23	0.913043	NC
R05	27157	G	A	59.694	330	0.972727	Ser→Asn
B3	27499	T	C	55.257	12	0.916667	Ser→Pro
R15	27679	C	T	69.237	68	0.985294	Leu→Phe
E5	28077	G	C	57.776	317	0.930599	Val→Leu
R15	28077	G	C	56.227	74	0.972973	Val→Leu
R07	28084	A	G	55.62	26	0.961538	Glu→Gly
R01	28144	T	C	52.609	11	0.909091	Leu→Ser
G1	28144	T	C	56.015	2373	0.91909	Leu→Ser
A2	28144	T	C	40.532	50	0.86	Leu→Ser
C2	28144	T	C	54.613	2824	0.923159	Leu→Ser
A4	28144	T	C	55.813	1083	0.925208	Leu→Ser
E5	28144	T	C	53.439	319	0.909091	Leu→Ser
G11	28144	T	C	57.476	3389	0.944821	Leu→Ser
R15	28144	T	C	64.466	74	0.972973	Leu→Ser
R15	28391	C	T	51.175	74	0.959459	Arg→Cys
C9	28514	G	T	61.455	59	0.898305	Asp→Tyr
R06	28688	T	C	49.151	78	0.897436	NC
C12	28922	G	T	64.466	84	0.880952	Ala→Ser
C2	29095	C	T	54.923	2349	0.941252	NC
C1	29483	T	G	42.187	318	0.811321	Ser→Ala

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**NANO · MICRO
small**

www.small-journal.com

Figure 7. Genome wide SNP analysis. a) Allele frequency of 50 NTS positive samples (top) and 1145 samples from GISAID (bottom). The altered nucleotides are colored with four different colors. We removed samples with uncertain nucleotides when calculating allele frequency of the 50 NTS positive samples for each SNP (Experimental Section). b) A part of the LD plot contained the most high *D* value mutation pairs. Mutations G28077C and C24034T have the highest *D* value. c) Typing result of 31 NTS positive samples.

sample contamination. Meanwhile, integrated with automated or semiautomated platforms to reduce the manual operation and improve the detection throughput, sequencing data analyzed by cloud analysis may also be introduced for quick high-throughput detection.

4. Experimental Section

Primer Panel Design for SARS-CoV-2: The SARS-CoV-2 primer panel was designed to simultaneously detect virus virulence- and infection-related genes and variants thereof. The 21563–29674 bp genome region, containing the genes encoding S, ORF3a, E, M, ORF6, ORF7a, ORF8, N, and ORF10, was selected as a template to design a series of end-to-end primers. The region encoding ORF1ab was selected as a template to design a nested primer for the higher sensitivity detection of SARS-CoV-2. All primers were designed using the online tool Primer-BLAST (https://www.ncbi.nlm.nih.c/tools/primer-blast/), and the specificity of all primers was verified against *Homo sapiens*, fungi, and bacteria. Finally, *N*, *S*, *rdrp*, and *E* gene sequences of SARS-related viruses available at GenBank were downloaded and selected on January 1, 2020 (accession NC_045512). Multiple sequence alignment of SARS-CoV-2 against SARS-related viruses was performed using Clustal W (version 1.83) for each gene individually, and the alignment was used for the in silico evaluation of the specificity of the designed primers to SARS-CoV-2. All specific primers were collected to form the SARS-CoV-2 primer panel.

Primer Panel Design for the Detection of Ten Kinds of Respiratory Virus: The target genes for each virus were selected based on previous

literature, and all complete and partial gene sequences available in GenBank through November 1, 2019, were downloaded. The list for each target gene was manually checked and artificial sequences (e.g., lab-derived, synthetic), along with sequence duplicates, were removed, resulting in a final list. Multiple sequence alignment was performed using Clustal W (version 1.83) for each gene individually, and the variation rate of each base was calculated using an in-house pipeline. The final primers for each virus were manually selected following the previous metrics[39] for multiplex PCR design, with an expected amplicon length, ranging from 300 to 800 bp.

NTS Detection Method: The total nucleic acid of sample was pretreated using the PrimeScript II 1st Strand cDNA Synthesis Kit (Takara Bio, Japan) in a 10 µL reaction system with 1 µL random six hexamers (50×10^{-6} M), 1 µL dNTP mixture, and 8 µL total nucleic acid at 95 °C for 5 min and 4 °C for 2 min. The cDNA was synthesized using the PrimeScript II 1st Strand cDNA Synthesis Kit (Takara Bio, Japan) in a 20 µL reaction system with 10 µL pretreated total nucleic acid, 0.5 µL RNase Inhibitor, 4 µL 5× PrimeScript II Buffer, 1 µL PrimeScript II Rtase, and 4.5 µL RNase Free ddH$_2$O. The product of synthesized cDNA was then purified with 1 × AMpure beads (Beckman Coulter, USA) and eluted in 10 µL TE buffer. The target genes were amplified using the SARS-CoV-2 or respiratory virus primer panel in a 20 µL reaction system with 5 µL eluate, 5 µL primer (10×10^{-6} M), and 10 µL 2 × Phusion U Multiplex PCR Master Mix (Thermo Fisher, USA) following previous research with several modification.[40,41] Multiplex amplification was performed in a C1000 Thermocycler (Bio-Rad, USA) using the following procedure: 1 cycle at 94 °C for 3 min and 30 cycles at 95 °C for 10 s, 55 °C for 30 s, and 68 °C for 20 s, followed by a final elongation step at 68 °C for 5 min. The product of the first-step was purified with

Table 2. Results of NTS for detecting SARS-CoV-2 and common respiratory viruses. ND: not detected.

Patients ID	NTS result (SARS-CoV-2)	NTS result (common respiratory viruses)
F11	Positive	ND
E11	Positive	human influenza A virus H3N2
A11	Positive	ND
B9	Positive	ND
C9	Positive	ND
D9	Positive	ND
D11	Positive	ND
C12	Positive	ND
E6	Positive	ND
B3	Positive	ND
E12	Positive	ND
G6	Inconclusive	ND
B1	Positive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Positive control	Positive	ND
Negative control	ND	ND
Negative control	ND	ND

0.8 × AMpure beads (Beckman Coulter, USA) and eluted in 10 µL TE buffer. Then, 5 µL eluate was used for barcoding PCR, with 5 µL barcoded SARS-CoV-2 or respiratory virus primer panel (10 × 10^{-6} M) and 10 µL 2 × Phusion U Multiplex PCR Master Mix, using the following procedure: 1 cycle at 94 °C for 3 min and 15 cycles at 95 °C for 10 s, 55 °C for 30 s, and 68 °C for 20 s. The barcode sequence was from the Nanopore PCR barcode kit EXP-PBC096 (Oxford nanopore technologies, UK), and all primer oligos and full-length *S* and *N* gene fragments were synthesized by Genscript (China). Equal masses of the products of barcoding PCR from the different samples were pooled. TE buffer was assayed in each batch as a negative control. Sequencing libraries were constructed using the 1D Ligation Kit (SQK-LSK109; Oxford Nanopore, UK) and sequenced using Oxford Nanopore MinION or GridION.

LoD of the NTS Test: The NTS library was prepared from a virus-negative nasopharyngeal swab spiked with plasmids containing synthetic *S* and *N* genes of COVID-19 at concentrations of 0, 10, 100, 500, 1000, and 3000 copies per reaction, with four replicates at each concentration. The NTS libraries were prepared as described above and sequenced using MinION for 10, 30 min, 1, 2, and 4 h. The sequencing data were processed as described for virus identification. The LoD was determined when the concentration of reads mapped to COVID-19 was significantly higher than that mapped for the negative control in 3/4 replicates.

Nanopore Sequencing Data Processing: Basecalling and quality assessment for MinION sequencing data were performed using high accuracy mode in the Guppy (v. 3.1.5) software; for GridION, the process was conducted using MinKNOW (v. 3.6.5) integrated in the instrument. Sequencing reads with low quality (*Q* score < 7, filter to "fail" by MinKNOW) and undesired length (<200 nt) were discarded. This quality control cut-off ensured that the mean sequencing accuracy was above 85% in accordance with that reported in other studies on ONT nanopore sequencing.[15,36,42,43] Next, Porechop (v. 0.2.4)[44] was used for adaptor trimming and barcode demultiplexing for retained reads with the parameter-barcode_threshold 85.

Mapping Tool and Mapping Database: BLASTn (v. 2.9.0+)[45] was used to map the reads of each sample against the virus genome reference database. The blast parameters were set as identity ≥90% and *E* value = 1e−05. All virus genomic sequences were downloaded on January 20, 2020, from NCBI Refseq FTP (http://ftp.ncbi.nlm.nih.gov/refseq/release/viral/); the SARS-CoV-2 genome was added to the BLAST database because the genome was not collected in the virus Refseq database prior to January 20, 2020. The taxonomy of each read was assigned according to the taxonomic information of the mapped subject sequence.

Reads quality and mapping identity evaluation: Sequencing reads were mapped to Human (GRCh38 version) and SARS-CoV-2 (NC_045512.2) genomes with minimap2 (v2.17-r941).[46] The average reads quality and alignment identity was calculated by NanoPlot (v1.29.1).[47]

Interpretation of NTS Results: After sequencing, the amplified fragments were mapped with all virus genomic sequences. If the highest identity of reads mapped was non-SARS-CoV-2 virus, then the virus with the highest identity was judged as the result. If the highest identity of reads mapped to the SARS-CoV-2 virus, then the judgment will perform the NTS score judgment. In detail, the sequencing data were obtained at regular intervals after sequencing and then filtered (identity ≥90%) to obtain valid reads. For determining whether the target was SARS-CoV-2 positive, interpretation was performed using the previous rule, with modification.[32–34] In brief, if the read matched a region within 50 bp upstream of the start and 50 bp downstream of the end of the design fragment, the read was counted. The mapping score was determined as 1, 0.4, or 0 when the ratio of count number in the sample to that in the negative control of each target was >10, between 3 and 10, or <3. The total mapping score of each target was summed and samples with >2.4 total mapping score were defined as positive for SARS-CoV-2 infection; 1.2 to 2.4 total mapping score indicated an inconclusive result, and <1.2 total mapping score was considered to indicate negative for infection. For determination of the other ten kinds of common respiratory virus, a sample was considered positive for the virus if it was positive for at least one designed site, otherwise it was negative.

Sample Collection: Throat swab samples were collected by healthcare workers based on clinical indications. Samples were collected in 10 mL of Viral Transport Medium (Becton Dickinson, USA) and transported to a clinical laboratory where they were processed immediately. Swabs were vortexed in 1 mL of TE buffer and centrifuged at 20000 × g for 10 min. The supernatant was removed and 200 µL of the specimen was retained for total nucleic acid extraction, which was performed using 200 µL of pretreated samples using the Sansure SUPRall DNA Extraction Kit (Changsha, China), following the manufacturer's instructions. Extracted total nucleic acid was stored at −70 °C until RT-qPCR or NTS testing.

Samples were selected for inclusion in this study based on two criteria:

1) Random selection of selected 45 suspected COVID-19 outpatients who had undergone RT-qPCR testing; the RT-qPCR-tested nucleic acid was stored at –70 °C to allow for NTS testing.

2) Random selection of 16 hospitalized patients with COVID-19 at Renmin Hospital of Wuhan University in Wuhan during February 2020; the RT-qPCR-tested nucleic acid was stored at –70 °C for further NTS testing. All hospitalized patients were diagnosed with COVID-19 infection. These samples were used to confirm whether NTS can detect samples that are mistaken for false negatives by RT-qPCR.

RT-qPCR for Confirmation of SARS-Cov-2 Infection: The total isolated nucleic acid was used for RT-qPCR assaying following the manufacturer's instructions. Briefly, RT-qPCR was carried out in a 25 µL reaction system using a novel coronavirus RT-qPCR kit (kit 1, Huirui, China) with 5 µL total nucleic acid or a 20 µL reaction system using the 2019-nCoV RT-qPCR kit (kit 2, BioGerm, China) with 5 µL total nucleic acid. For kit 1, amplification was performed using a Quantstudio Dx Real-time PCR system (Thermo Fisher, USA) using the following procedure: 1 cycle at 50 °C for 15 min and 95 °C for 5 min and 35 cycles at 95 °C for 10 s and 55 °C for 40 s. The FAM and ROX fluorescence channels were used to

detect *orf1ab* and *N*, respectively. Successful amplification of both genes and a Ct value ≤35 was recognized as positive for SARS-CoV-2 infection; a Ct value between 35.2 and 39.2 was recognized as inconclusive, and one of the genes being undetected or a Ct ≥ 39.2 was recognized as negative. For kit 2, amplification was performed in a CFX96 Thermocycler (Bio-Rad) using the following procedure: 1 cycle at 50 °C for 10 min and 95 °C for 5 min and 35 cycles at 95 °C for 10 s and 55 °C for 40 s. The FAM, HEX, and CY5 fluorescence channels were used to detect *orf1ab*, *E*, and *N*, respectively. This kit only utilized the results of the *orf1ab* and *N* gene to reach a conclusion. Successful amplification of both genes and a Ct value ≤35 was recognized as positive for SARS-CoV-2 infection; only one site with a Ct value ≤35 or both genes having Ct values between 35 and 38 was recognized as inconclusive, and no successful amplification or Ct ≥ 38 was recognized as negative for infection.

Candidate Variant Calling: Medaka (v0.11.5),[48] an analysis tool developed by Oxford Nanopore Technologies that uses a neural network algorithm, was used for calling variants. The performance of Medaka was confirmed by Gilpatrick et al.[49] Variant calling for haploid genomes contained two steps. First, the reads of SARS-CoV-2 were aligned to the reference genome of NC_045512.2 using minimap2,[46] and medaka_consensus generated probable consensus sequences using the trained model r941_min_high_g303. Then the variants were called by the medaka_variant program based on the consensus sequences and reference genome. Due to the difficulties of detecting indel variants from nanopore sequencing data, only the candidate variants to single nucleotide substitutions were considered. The variants within certainty regions with at least 10 × sequencing depth and with output quality score over 30 were accepted as candidate nucleotide mutations.

LS Type Classification: For each of the 50 NTS positive samples, the fraction of reads supporting cytosine and thymine at 28144 was calculated. A sample was assigned to S type if the depth (at the 28144 locus) was ≥10 and cytosine fraction was ≥75%; a sample was assigned to L type if the depth was ≥10 and the thymine fraction was ≥75%; a sample was assigned to uncertain type if the depth was <10 or both nucleotide fractions were <75%.

Allele Frequency Calculation: For 42 single nucleotide mutations of the 50 NTS positive samples, the allele frequency calculations followed two steps. First, the samples containing uncertain nucleotides at the mutation locus were removed when calculating the allele frequency at this locus. The locus with depth <10 or the fraction of reads supporting the dominate nucleotide at the locus <75% was defined as an uncertain nucleotide locus (The coverages of A, T, C, G nucleotides were counted at the 42 mutation loci for each sample. The richest nucleotide was defined as the dominate nucleotide at the locus in each sample.). Then, the allele frequency of one mutation locus was calculated using the number of supporting samples divided by the number of samples with certain nucleotides at the mutation locus.

For 1145 SARS-CoV-2 genomes from GISAID, the genome sequences were first aligned to the NC_045512.2 reference sequence using minimap2. Every alternative nucleotide in the sample genomes were then identified, compared to the reference, and the alternative nucleotide fraction was calculated as the allele frequency at each locus.

Linkage Disequilibrium Analysis: R (3.6.3) package gaston (1.5.6) was used to build LD plots, which were designed for genotype SNP linkage disequilibrium analysis.[50] Therefore, a specific mutation matrix (27 samples, 32 unique mutation sites), as well as matrices containing samples information (names etc.) required by gaston was constructed. The *D* value was selected for LD plot, and values were calculated according to Equation (1)

$$D = P(AB) - P(A)P(B) \tag{1}$$

where $P(AB)$ is the frequency of mutations *A* and *B* co-occurring in the samples and $P(A)$ or $P(B)$ are the frequencies of mutations *A* or *B* in samples, respectively. Although the *D* value will have a high false-negative rate, it was sufficient to determine whether the mutations in the samples were significantly linked.

Clinical Records of Patients: The clinical records of patients were stored at Renmin Hospital of Wuhan University. Clinical, laboratory, and radiological characteristic data, as well as treatment history and outcome data were collected from electronic medical records. The data were reviewed by a trained team of physicians. The study and use of all records were approved by the Ethics committee of Renmin Hospital of Wuhan University (WDRY2019-056), consents from patients were waived by the Ethics committee.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

The authors declare no conflict of interest.

[1] W. J. Guan, Z. Y. Ni, Y. Hu, W. H. Liang, C. Q. Ou, J. X. He, L. Liu, H. Shan, C. L. Lei, D. S. C. Hui, B. Du, L. J. Li, G. Zeng, K. Y. Yuen, R. C. Chen, C. L. Tang, T. Wang, P. Y. Chen, J. Xiang, S. Y. Li, J. L. Wang, Z. J. Liang, Y. X. Peng, L. Wei, Y. Liu, Y. H. Hu, P. Peng, J. M. Wang, J. Y. Liu, Z. Chen, G. Li, Z. J. Zheng, S. Q. Qiu, J. Luo, C. J. Ye, S. Y. Zhu, N. S. Zhong, *N. Engl. J. Med.* **2020**, *382*, 1708.
[2] R. Liu, A. Fu, Z. Deng, Y. Li, T. Liu, *View* **2020**, *1*, 4.
[3] Y. Wang, H. Kang, X. Liu, Z. Tong, *J. Med. Virol.* **2020**, *92*, 538.
[4] A. A. Green, P. A. Silver, J. J. Collins, P. Yin, *Cell* **2014**, *159*, 925.
[5] K. Pardee, A. A. Green, T. Ferrante, D. E. Cameron, A. DaleyKeyser, P. Yin, J. J. Collins, *Cell* **2014**, *159*, 940.
[6] A. A. G. Pardee, M. K. Takahashi, D. Braff, G. Lambert, J. W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N. M. Daringer, I. Bosch, D. M. Dudley, D. H. O'Connor, L. Gehrke, J. J. Collins, *Cell* **2016**, *165*, 1255.
[7] A protocol for detection of COVID-19 using CRISPR diagnostics, https://www.broadinstitute.org/files/publications/special/COVID-19%20detection%20(updated).pdf (accessed: April 2020).
[8] G. J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, F. Zhang, *Science* **2017**, *356*, 438.

ADVANCED
SCIENCE NEWS

www.advancedsciencenews.com

NANO · MICRO
small

www.small-journal.com

[9] J. P. Broughton, X. Deng, G. Yu, C. L. Fasching, V. Servellita, J. Singh, X. Miao, J. A. Streithorst, A. Granados, A. Sotomayor-Gonzalez, K. Zorn, A. Gopez, E. Hsu, W. Gu, S. Miller, C. Y. Pan, H. Guevara, D. A. Wadford, J. S. Chen, C. Y. Chiu, *Nat. Biotechnol.* **2020**, https://doi.org/10.1038/s41587-020-0513-4.

[10] Y. C. Su, D. E. Anderson, B. E. Young, F. Zhu, M. Linster, S. Kalimuddin, J. G. Low, Z. Yan, J. Jayakumar, L. Sun, G. Z. Yan, I. H. Mendenhall, Y.-S. Leo, D. C. Lye, L.-F. Wang, G. J. Smith, *bioRxiv* **2020**, 2020.03.11.987222.

[11] M. R. Wilson, S. N. Naccache, E. Samayoa, M. Biagtan, H. Bashir, G. Yu, S. M. Salamat, S. Somasekar, S. Federman, S. Miller, R. Sokolic, E. Garabedian, F. Candotti, R. H. Buckley, K. D. Reed, T. L. Meyer, C. M. Seroogy, R. Galloway, S. L. Henderson, J. E. Gern, J. L. DeRisi, C. Y. Chiu, *N. Engl. J. Med.* **2014**, *370*, 2408.

[12] H. C. Metsky, C. B. Matranga, S. Wohl, S. F. Schaffner, C. A. Freije, S. M. Winnicki, K. West, J. Qu, M. L. Baniecki, A. Gladden-Young, A. E. Lin, C. H. Tomkins-Tinch, S. H. Ye, D. J. Park, C. Y. Luo, K. G. Barnes, R. R. Shah, B. Chak, G. Barbosa-Lima, E. Delatorre, Y. R. Vieira, L. M. Paul, A. L. Tan, C. M. Barcellona, M. C. Porcelli, C. Vasquez, A. C. Cannons, M. R. Cone, K. N. Hogan, E. W. Kopp, J. J. Anzinger, K. F. Garcia, L. A. Parham, R. M. G. Ramirez, M. C. M. Montoya, D. P. Rojas, C. M. Brown, S. Hennigan, B. Sabina, S. Scotland, K. Gangavarapu, N. D. Grubaugh, G. Oliveira, R. Robles-Sikisaka, A. Rambaut, L. Gehrke, S. Smole, M. E. Halloran, L. Villar, S. Mattar, I. Lorenzana, J. Cerbino-Neto, C. Valim, W. Degrave, P. T. Bozza, A. Gnirke, K. G. Andersen, S. Isern, S. F. Michael, F. A. Bozza, T. M. L. Souza, I. Bosch, N. L. Yozwiak, B. L. MacInnis, P. C. Sabeti, *Nature* **2017**, *546*, 411.

[13] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G. Xiao, Z.-L. Shi, *Nature* **2020**, *579*, 270.

[14] D. Deamer, M. Akeson, D. Branton, *Nat. Biotechnol.* **2016**, *34*, 518.

[15] T. Charalampous, G. L. Kay, H. Richardson, A. Aydin, R. Baldan, C. Jeanes, S. Rae, S. Grundy, D. J. Turner, J. Wain, R. M. Leggett, D. M. Livermore, J. O'Grady, *Nat. Biotechnol.* **2019**, *37*, 783.

[16] Y. Xu, K. Lewandowski, K. Jeffery, L. O. Downs, D. Foster, N. D. Sanderson, J. Kavanagh, A. Vaughan, C. Salvagno, R. Vipond, M. Carroll, R. Danby, T. Peto, D. Crook, A. S. Walker, P. C. Matthews, S. T. Pullan, *J. Infect.* **2020**, *80*, 571.

[17] K. Lewandowski, Y. Xu, S. T. Pullan, S. F. Lumley, D. Foster, N. Sanderson, A. Vaughan, M. Morgan, N. Bright, J. Kavanagh, R. Vipond, M. Carroll, A. C. Marriott, K. E. Gooch, M. Andersson, K. Jeffery, T. E. A. Peto, D. W. Crook, A. S. Walker, P. C. Matthews, *J. Clin. Microbiol.* **2019**, *58*, e00963.

[18] G. Taiaroa, D. Rawlinson, L. Featherstone, M. Pitt; L. Caly; J. Druce; D. Purcell; L. Harty; T. Tran; J. Roberts; N. Scott; M. Catton; D. Williamson, L. Coin, S. Duchene, *bioRxiv* **2020**, 2020.03.05.976167.

[19] D. Kim, J. Y. Lee, J. S. Yang, J. W. Kim, V. N. Kim, H. Chang, *Cell* **2020**, *181*, 914.

[20] A. Viehweger, S. Krautwurst, K. Lamkiewicz, R. Madhugiri, J. Ziebuhr, M. Holzer, M. Marz, *Genome Res.* **2019**, *29*, 1545.

[21] N. R. Faria, J. Quick, I. M. Claro, J. Theze, J. G. de Jesus, M. Giovanetti, M. U. G. Kraemer, S. C. Hill, A. Black, A. C. da Costa, L. C. Franco, S. P. Silva, C. H. Wu, J. Raghwani, S. Cauchemez, L. du Plessis, M. P. Verotti, W. K. de Oliveira, E. H. Carmo, G. E. Coelho, A. Santelli, L. C. Vinhal, C. M. Henriques, J. T. Simpson, M. Loose, K. G. Andersen, N. D. Grubaugh, S. Somasekar, C. Y. Chiu, J. E. Munoz-Medina, C. R. Gonzalez-Bonilla, C. F. Arias, L. L. Lewis-Ximenez, S. A. Baylis, A. O. Chieppe, S. F. Aguiar, C. A. Fernandes, P. S. Lemos, B. L. S. Nascimento, H. A. O. Monteiro, I. C. Siqueira, M. G. de Queiroz, T. R. de Souza, J. F. Bezerra, M. R. Lemos, G. F. Pereira, D. Loudal, L. C. Moura, R. Dhalia, R. F. Franca, T. Magalhaes, E. T. Marques Jr., T. Jaenisch, G. L. Wallau, M. C. de Lima, V. Nascimento, E. M. de Cerqueira, M. M. de Lima, D. L. Mascarenhas, J. P. M. Neto, A. S. Levin, T. R. Tozetto-Mendoza, S. N. Fonseca, M. C. Mendes-Correa, F. P. Milagres, A. Segurado, E. C. Holmes, A. Rambaut, T. Bedford, M. R. Nunes, E. C. Sabino, L. C. J. Alcantara, N. J. Loman, O. G. Pybus, *Nature* **2017**, *546*, 406.

[22] J. Quick, N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouedraogo, B. Afrough, A. Bah, J. H. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, A. Camino-Sanchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallash, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, R. Trina, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O'Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Woelfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keita, A. Rambaut, P. Formenty, S. Gunther, M. W. Carroll, *Nature* **2016**, *530*, 228.

[23] A. Harcourt, A. Tamin, X. Lu, S. Kamili, S. K. Sakthivel, L. Wang, J. Murray, K. Queen, B. Lynch, B. Whitaker, Y. Tao, C. R. Paden, J. Zhang, Y. Li, A. Uehara, H. Wang, C. Goldsmith, H. A. Bullock, R. Gautam, C. Schindewolf, K. G. Lokugamage, D. Scharton, J. A. Plante, D. Mirchandani, S. G. Widen, K. Narayanan, S. Makino, T. G. Ksiazek, K. S. Plante, S. C. Weaver, S. Lindstrom, S. Tong, V. D. Menachery, N. J. Thornburg, *bioRxiv* **2020**, 2020.03.02.972935.

[24] J. F. Chan, S. Yuan, K. H. Kok, K. K. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C. Yip, R. W. Poon, H. W. Tsoi, S. K. Lo, K. H. Chan, V. K. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. Cheng, H. Chen, C. K. Hui, K. Y. Yuen, *Lancet* **2020**, *395*, 514.

[25] K. K. To, O. T. Tsang, W. S. Leung, A. R. Tam, T. C. Wu, D. C. Lung, C. C. Yip, J. P. Cai, J. M. Chan, T. S. Chik, D. P. Lau, C. Y. Choi, L. L. Chen, W. M. Chan, K. H. Chan, J. D. Ip, A. C. Ng, R. W. Poon, C. T. Luo, V. C. Cheng, J. F. Chan, I. F. Hung, Z. Chen, H. Chen, K. Y. Yuen, *Lancet Infect. Dis.* **2020**, *20*, 565.

[26] S. C. Moore, R. Penrice-Randal, M. Alruwaili, X. Dong, S. T. Pullan, D. Carter, K. Bewley, Q. Zhao, Y. Sun, C. Hartley, E.-m. Zhou, T. Solomon, M. B. J. Beadsworth, J. Cruise, D. Bogaert, D. W. T. Crook, D. A. Matthews, A. D. Davidson, Z. Mahmood, W. Aljabr, J. Druce, R. T. Vipond, L. Ng, L. Renia, P. Openshaw, J. K. Baillie, M. W. Carroll, C. Semple, L. Turtle, J. A. Hiscox, *medRxiv* **2020**, 2020.03.05.20032011.

[27] J. R. Fauver, M. E. Petrone, E. B. Hodcroft, K. Shioda, H. Y. Ehrlich, A. G. Watts, C. B. F. Vogels, A. F. Brito, T. Alpert, A. Muyombwe, J. Razeq, R. Downing, N. R. Cheemarla, A. L. Wyllie, C. C. Kalinich, I. M. Ott, J. Quick, N. J. Loman, K. M. Neugebauer, A. L. Greninger, K. R. Jerome, P. Roychoudhury, H. Xie, L. Shrestha, M.-L. Huang, V. E. Pitzer, A. Iwasaki, S. B. Omer, K. Khan, I. I. Bogoch, R. A. Martinello, E. F. Foxman, M. L. Landry, R. A. Neher, A. I. Ko, N. D. Grubaugh, *Cell* **2020**, *181*, 990.

[28] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan,

ADVANCED
SCIENCE NEWS

www.advancedsciencenews.com

NANO · MICRO

small

www.small-journal.com

Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, *Lancet* **2020**, *395*, 565.

[29] J. Lu, L. du Plessis, Z. Liu, V. Hill, M. Kang, H. Lin, J. Sun, S. Francois, M. U. G. Kraemer, N. R. Faria, J. T. McCrone, J. Peng, Q. Xiong, R. Yuan, L. Zeng, P. Zhou, C. Liang, L. Yi, J. Liu, J. Xiao, J. Hu, T. Liu, W. Ma, W. Li, J. Su, H. Zheng, B. Peng, S. Fang, W. Su, K. Li, R. Sun, R. Bai, X. Tang, M. Liang, J. Quick, T. Song, A. Rambaut, N. Loman, J. Raghwani, O. G. Pybus, C. Ke, *Cell* **2020**, *181*, 997.

[30] M. Wang, Q. Wu, W. Xu, B. Qiao, J. Wang, H. Zheng, S. Jiang, J. Mei, Z. Wu, Y. Deng, F. Zhou, W. Wu, Y. Zhang, Z. Lv, J. Huang, X. Guo, L. Feng, Z. Xia, D. Li, Z. Xu, T. Liu, P. Zhang, Y. Tong, Y. Li, *medRxiv* **2020**, 2020.02.12.20022327.

[31] M. Xiao, X. Liu, J. Ji, M. Li, J. Li, L. Yang, W. Sun, P. Ren, G. Yang, J. Zhao, T. Liang, H. Ren, T. Chen, H. Zhong, W. Song, Y. Wang, Z. Deng, Y. Zhao, Z. Ou, D. Wang, J. Cai, X. Cheng, T. Feng, H. Wu, Y. Gong, H. Yang, J. Wang, X. Xu, S. Zhu, F. Chen, Y. Zhang, W. Chen, Y. Li, J. Li, *bioRxiv* **2020**, 2020.03.16.993584.

[32] T. A. Blauwkamp, S. Thair, M. J. Rosen, L. Blair, M. S. Lindner, I. D. Vilfan, T. Kawli, F. C. Christians, S. Venkatasubrahmanyam, G. D. Wall, A. Cheung, Z. N. Rogers, G. Meshulam-Simon, L. Huijse, S. Balakrishnan, J. V. Quinn, D. Hollemon, D. K. Hong, M. L. Vaughn, M. Kertesz, S. Bercovici, J. C. Wilber, S. Yang, *Nat. Microbiol.* **2019**, *4*, 663.

[33] C. Langelier, M. S. Zinter, K. Kalantar, G. A. Yanik, S. Christenson, B. O'Donovan, C. White, M. Wilson, A. Sapru, C. C. Dvorak, S. Miller, C. Y. Chiu, J. L. DeRisi, *Am J. Respir. Crit. Care Med.* **2018**, *197*, 524.

[34] H. C. Metsky, K. J. Siddle, A. Gladden-Young, J. Qu, D. K. Yang, P. Brehio, A. Goldfarb, A. Piantadosi, S. Wohl, A. Carter, A. E. Lin, K. G. Barnes, D. C. Tully, B. Corleis, S. Hennigan, G. Barbosa-Lima, Y. R. Vieira, L. M. Paul, A. L. Tan, K. F. Garcia, L. A. Parham, I. Odia, P. Eromon, O. A. Folarin, A. Goba, C. Viral Hemorrhagic Fever, E. Simon-Loriere, L. Hensley, A. Balmaseda, E. Harris, D. S. Kwon, T. M. Allen, J. A. Runstadler, S. Smole, F. A. Bozza, T. M. L. Souza, S. Isern, S. F. Michael, I. Lorenzana, L. Gehrke, I. Bosch, G. Ebel, D. S. Grant, C. T. Happi, D. J. Park, A. Gnirke, P. C. Sabeti, C. B. Matranga, *Nat. Biotechnol.* **2019**, *37*, 160.

[35] S. Wei, Z. R. Weiss, Z. Williams, *G3: Genes, Genomes, Genet.* **2018**, *8*, 1649.

[36] B. D. Wilson, M. Eisenstein, H. T. Soh, *Anal. Chem.* **2019**, *91*, 10204.

[37] X. Tang, C. Wu, X. Li, Y. Song, X. Yao, X. Wu, Y. Duan, H. Zhang, Y. Wang, Z. Qian, J. Cui, J. Lu, *Natl. Sci. Rev.* **2020**, *0*, 1.

[38] GISAID database, https://www.gisaid.org/, (accessed: March 2020).

[39] Z. Shen, W. Qu, W. Wang, Y. Lu, Y. Wu, Z. Li, X. Hang, X. Wang, D. Zhao, C. Zhang, *BMC Bioinformatics* **2010**, *11*, 143.

[40] J. Quick, N. D. Grubaugh, S. T. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. G. de Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara Jr., E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen, N. J. Loman, *Nat. Protoc.* **2017**, *12*, 1261.

[41] N. D. Grubaugh, K. Gangavarapu, J. Quick, N. L. Matteson, J. G. De Jesus, B. J. Main, A. L. Tan, L. M. Paul, D. E. Brackney, S. Grewal, N. Gurfield, K. K. A. Van Rompay, S. Isern, S. F. Michael, L. L. Coffey, N. J. Loman, K. G. Andersen, *Genome Biol.* **2019**, *20*, 8.

[42] R. M. Leggett, C. Alcon-Giner, D. Heavens, S. Caim, T. C. Brook, M. Kujawska, S. Martin, N. Peel, H. Acford-Palmer, L. Hoyles, P. Clarke, L. J. Hall, M. D. Clark, *Nat. Microbiol.* **2020**, *5*, 430.

[43] M. Jain, I. T. Fiddes, K. H. Miga, H. E. Olsen, B. Paten, M. Akeson, *Nat. Methods* **2015**, *12*, 351.

[44] Porechop: adapter trimmer for Oxford Nanopore reads, https://github.com/rrwick/Porechop (accessed: January 2020).

[45] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403.

[46] H. Li, *Bioinformatics* **2018**, *34*, 3094.

[47] W. De Coster, S. D'Hert, D. T. Schultz, M. Cruts, C. Van Broeckhoven, *Bioinformatics* **2018**, *34*, 2666.

[48] Medaka: sequence correction provided by ONT Research, https://github.com/nanoporetech/medaka (accessed: January 2020).

[49] T. Gilpatrick, I. Lee, J. E. Graham, E. Raimondeau, R. Bowen, A. Heron, B. Downs, S. Sukumar, F. J. Sedlazeck, W. Timp, *Nat. Biotechnol.* **2020**, *38*, 433.

[50] Gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models, https://CRAN.R-project.org/package=gaston (accessed: March 2020).

[51] Specific primers and probes for detection 2019 novel coronavirus (recommend by Chinese CDC), http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html (accessed: February 2020).

[52] Research Use Only 2019-Novel Coronavirus (2019-nCoV) Real-time RT-PCR Primers and Probes (recommend by American CDC), https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html (accessed: February 2020).

[53] V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, T. Bleicker, S. Brunink, J. Schneider, M. L. Schmidt, D. G. Mulders, B. L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J. L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M. P. Koopmans, C. Drosten, *Eurosurveillance* **2020**, *25*, 2000045.